

# Supporting Information

Doyle and Csete 10.1073/pnas.1103557108

## SI Text

This supporting information has some very elementary tutorial introductions to minimal control theory, with details that are easily checked and experiments that are easily performed. The larger research program that we are pursuing can be summarized as if you accept the empirical and theoretical claims in neuroscience, then what is the right mathematical framework to connect the mechanistic details with the behavior? Tremendous theoretical progress is being made on what we believe is the right combination of robust, layered, distributed, and nonlinear control, integrated as needed with communication and computation theories. At best, the state of the art is nascent and promising but technically inaccessible, and it has been focused on computer and energy networks, not neuroscience. Because the core theory is being developed and vetted among control theorists, we will aim to build language (not standardized or well-thought out so far) that can be used to connect the theory with real behavior and known constraints on the components. Essential to this dialogue is a shared understanding of layered architectures, robustness, feedback, dynamics, and optimization as well as why a constraints-based theory is both theoretically and biologically natural. What follows are simple examples that complement the layering ideas in the text.

## Minimal Control Theory

The amplifier circuit in Fig. S1 is the simplest feedback system possible, and it will be used as a minimal starting point to start formalizing aspects of the constraints view of architecture motivated by the case studies in the paper. Then, we will briefly sketch out the theory underlying the pendulum case study, which is much more technical but still uses only undergraduate mathematics.

There are amplifiers loosely based on the principle described below all around us. Most familiar might be the audio amplifiers and tuning knobs for radios and audio equipment. The large/thin element is that, say, a high-tech radio has many knobs to adjust to get the station, volume, compensation for room acoustics, etc. Thus, there is a nearly infinite range of functional parameter values for the radio that can be explicitly tuned with just a few knobs. However, if you break open the box, there is a much larger number of hidden parameters whose variation can destroy the function inside. These variations are not allowed in normal use, and this use would be breaking the system. Even minor but random rewiring of the internal connections will ruin the radio. Also, if you make a math model of the circuit, the internal parameters are infinitely more numerous than the external tunable ones, and the math model will have the property that the set of functional parameters is both vastly large in absolute terms and also vanishingly thin as a fraction of all possible (random) perturbations.

Fig. S1 is the simplest possible math model that illustrates roughly how real feedback amplifiers work. The signals are an external input  $r$  and output  $z$ , an amplifier noise  $n$ , internal measurement  $y$ , and control  $u$ . The components are an (open-loop) amplifier  $A$  and controller  $C$ , giving the interconnected system output/input  $z/r$  gain  $G$  and noise response  $S$ . All of these quantities are assumed to be (possibly uncertain) real numbers, which capture the small signal (linear) steady state features of the circuit. It is the interplay between feedback and dynamics that is, ultimately, of most interest but also most confusing; therefore, we will first explore feedback without dynamics as a small step.

The basic properties of this circuit can all be derived using elementary algebra, including solving for the relationship among the external signals  $z = Gr + Sn$  as (Eq. S1)

$$\begin{aligned} z &= A(r + u) + n = A(r + Cz) + n \\ \Rightarrow z &= \left( \frac{1}{1 - AC} \right) (Ar + n) \triangleq Gr + Sn. \end{aligned} \quad [\text{S1}]$$

Powerful but precise amplifiers seem to be essential throughout technology and biology and are, in any case, ubiquitous. What is typically desired in electrical, mechanical, and/or hydraulic technologies as well as in control muscles ( $A$ ) with nerves ( $C$ ) is that the gain  $G$  be large and precise and the noise response  $S$  small. This model is an extremely abstracted and minimal model of such systems.

If there are no constraints and no uncertainty (i.e.,  $A$  and  $C$  can be specified arbitrarily and exactly and the noise  $n = 0$ ), then feedback is unnecessary, and any gain  $G$  can be realized perfectly with amplifier  $A = G$  and  $C = 0$ . Real high-gain amplifiers unavoidably have both noise and limits on their gains, and therefore, feedback is necessary; there are also limits on the achievable closed-loop performance and robustness that we will explore. There are, again, three distinct but interrelated types of constraints: (i) component constraints, (ii) system constraint, and (iii) protocol constraints. Mathematically, a minimal set of components constraints would be (S2)

$$\text{Component: } A \geq A_{\min}, 0 < -C < 1, n \neq 0. \quad [\text{S2}]$$

Typically, the noise  $n$  would have more characterization, but for our purposes here, its presence or absence is all that will be considered. Also,  $A \geq A_{\min} \gg 1$  would be a high but uncertain amplifier gain, with only the lower-bound  $A_{\min}$  assumed to be known (and thus, an upper bound on  $A$  is not assumed), whereas  $C$  is a (negative) feedback that has small gain, which can be tuned exactly to any desired value  $0 < -C < 1$ . In practice,  $C$  might be implemented with a potentiometer to give a (precisely) variable gain amplifier, and this finding is roughly what a volume knob on a radio does. These are typical tradeoffs that appear in real elementary components that can have either high or precise gains but not both. In all cases, larger  $A_{\min}$  would be more costly (e.g., larger and/or require a larger power supply), and therefore, the choice of  $A_{\min}$  would be a key design decision involving tradeoffs between these factors. The amplifier is also assumed to have noise  $n \neq 0$ , which will force constraints on  $S$ .

The corresponding minimal constraints on the system can be expressed as (Eq. S3)

$$\text{System: } S \leq S_{\max}, G \in [G_{\min}, G_{\max}] = G_{\max}[(1 - S_{\max}), 1]. \quad [\text{S3}]$$

What is typically required is to have moderately large and bounded gain  $G \in [G_{\min}, G_{\max}]$ , with small gain  $S \leq S_{\max}$  on the noise  $n$ . Note that, in this simplified formulation,  $S$  (called the sensitivity function in control theory) measures both the effects of noise on the output and the uncertainty in the gain  $G \in [G_{\min}, G_{\max}] = G_{\max}[(1 - S_{\max}), 1]$ , whereas in general, these numbers could be different. Robust systems would have  $S_{\max} \ll 1$ , which rejects noise and has small gain error, and high-performance systems would also have  $G_{\min} \gg 1$ . The protocol constraints, which are described algebraically as (Eq. S4)

$$\text{Protocol: } G = AS, S = \frac{1}{1-AC}, \quad [\text{S4}]$$

are depicted schematically in Fig. S1 (deliberately drawn to emphasize that lower and higher are purely conventional and have no intrinsic meaning).

If we start with the component and protocol constraints in expression S2 and Eq. S4, we can easily derive the resulting system constraints as (Eq. S5)

$$S \leq S_{\max} = \frac{1}{1-A_{\min}C}, G \in [G_{\min}, G_{\max}] = \frac{1}{-C}[(1-S_{\max}), 1]. \quad [\text{S5}]$$

Engineers call this process analysis, because it takes a given set of component constraints and their interconnections (from the protocols) and analyzes the resulting system properties, which are also naturally expressed as constraints. It is the obvious forward flow from components through protocols to systems described above for textiles and biology. In this context, the constraints in Eq. S5 are the consequences of those constraints in expression S2 and Eq. S4. However, we can also go the other way (backwards), starting with the system constraints in Eq. S3 and using the protocol constraints in Eq. S4 to synthesize or design the component constraints by solving for  $A$  and  $C$  to get (Eq. S6)

$$S \leq S_{\max} < 1, G \in [G_{\min}, G_{\max}], 1 < G_{\max}, G_{\min} = (1-S_{\max})G_{\max} \\ \Rightarrow C = \frac{-1}{G_{\max}} \Rightarrow A_{\min} \geq \frac{-1}{C} \left( \frac{1}{S_{\max}} - 1 \right). \quad [\text{S6}]$$

Engineers call this synthesis, because it involves starting with system requirements and architecture (i.e., protocol constraints) and synthesizing the necessary component constraints. Analysis can then be used to verify that this synthesis was successful. In this case, both directions are trivial, but synthesis is typically vastly harder and more complex than analysis, an issue not well-illustrated by this simplified example. Recall that this example is similar to the case studies in the paper where the feedbacks were far more complex than the more obvious forward assembly pathway. The iterative process of synthesis (usually using greatly simplified models and constraints) and analysis (with more complete models) contributes to system design, which in realistic situations, can involve many additional processes not considered here.

Where causality arises in Fig. S1 is in certain relationship between signals and systems. For example, the control input  $u$  can reasonably be said to be caused by the combination of controller  $C$  and signal  $y$  through  $u = Cy$ . A more subtle point that we are glossing over here is that the signals and systems in Fig. S1 are abstract objects, and they correspond to a functional decomposition of the system, not a physical one. That is, a physical circuit implementing Fig. S1 would not break up into the physical modules of controller and amplifier any more than a digital circuit is physically distinct from its analog implementation. Fortunately, this is a subject with abundant tutorial material, and therefore, the interested reader can easily find accessible explanations.

Beyond this trivial sense, the scientific jargon of causation and emergence provides little here, whereas the processes of analysis and synthesis are clearly defined after the three types of constraints are clarified. Neither controller nor amplifier causes the system behavior, and the protocols in this particular architecture imply a logical connection between component and systems constraints that can be used in either direction (i.e., analysis vs. synthesis). If we ask why there are the specific component constraints in expression S2, a proximal answer is that the underlying

technology makes it possible to fabricate components with these constraints and features. A more complete answer is that this technology also allows the systems constraints in Eq. S5 to be realized through the protocols in Eq. S4. If these constraints are not compatible, the architecture as specified is not viable, and much of engineering involves evaluating this possibility. Although it is possible to describe this idea in terms of up and down causation or emergence, these terms seem to add nothing to our understanding.

With these preliminaries, we can now briefly discuss design tradeoffs and parameter spaces. Fig. S2 plots the tradeoffs between system performance in terms of  $G_{\min}$ ,  $G_{\max}$ , and  $S_{\max}$  as a function of  $C$  for  $A_{\min} = 100$ . In Fig. S2 Right, note that, for positive feedback  $C > 0$ ,  $S_{\max} > 1$  and noise is amplified. Thus,  $C < 0$  is necessary and sufficient for  $S_{\max} < 1$ , a minimal robustness requirement. Given  $C < 0$ , then  $\frac{1}{A_{\min}} \ll |C|$  (equivalently  $|CA_{\min}| \gg 1$ ) is necessary and sufficient for  $S_{\max} \ll 1$ , in which case  $S_{\max} = |S_{\max}| \approx \left| \frac{1}{CA_{\min}} \right| \ll 1$ . However, for a given  $A_{\min}$ , there is a tradeoff between making  $|S_{\max}|$  small and  $G_{\min}$  large. Robust and functional amplifiers, thus, have both  $1 \ll A_{\min}$  and  $-1 \ll C \ll \frac{-1}{A_{\min}}$  to keep both  $|S_{\max}|$  small and  $G_{\min}$  large. The result is a tiny sliver of acceptable values of  $(C, A)$  in parameter space as shown in Fig. S3, which is nonetheless compatible with real engineering components.

This simple model shares key features with the textile and biological architectures sketched above. There is no great complexity here, but what little there is (i.e.,  $C \neq 0$ ) is needed only for robustness, not minimal functionality in an idealized setting with no uncertainty. The functional parameters are an infinitely large (deconstrained) set but a very small, thin (constrained) fraction of all possibly parameter values, and thus, random circuits are vanishingly unlikely to be robust or functional. What is missing completely is any notion of dynamics, the addition of which, in the final case study, adds enormously to the nature of the constraints involved. Also, the functional parameter sets in Fig. S3 are convex, which is emphatically not true in general. Indeed, reparameterizing problems to make them convex is the heart of research in robust, distributed, and nonlinear control. Controllers are almost never convex in natural coordinates. This trivial math model also illustrates that complexity is hidden in normal function. One cannot tell from the outside what exactly is going on inside unless one knows the architecture and probes it for robustness with perturbations. This finding seems true of biology as well.

### Simple Motor Control Example

For the last case study, we will consider a simple feedback motor control experiment that can easily be explored without specialized equipment, and it illustrates additional important constraints. The cartoon in Fig. S4 depicts the basic setup of a stabilization problem. The component constraints are that a mass  $m$  at location  $y$  on top of an inverted pendulum of length  $l$  is controlled by a muscle force  $u$  acting on a hand at position  $x$ , which is assumed to have effective mass  $M$ . The system constraint (and the main experimental problem) is that the hand must be controlled in such a way as to stabilize the up pendulum around  $\theta = 0$  using the eyes to see the location  $y$  of the pendulum's top. This experiment is a standard experiment in human motor control and undergraduate engineering, replacing the hand with an actuated cart, but for our purposes, it illustrates important concepts that can be formalized and made rigorous mathematically.

Simple variations in the constraints can yield radically different properties. If the pendulum is sufficiently long, then it is easy to learn to stabilize the up position, but if too short, it is impossible

for humans (although robots can be built that will outperform humans in this simple task). Eyes closed, sensing only through contact with the hand, is a more severe component constraint, and it is apparently impossible for humans to stabilize with any length pendulum. Finally, the down position is naturally stable and comparably easy to control. The easiest experimental demonstrations of these extreme differences are obtained with pendulums of widely varying lengths and tip masses, but they can also be performed with a standard mechanical extendable pointer.

These three cases are summarized in the cartoon in Fig. S5 in a way that can be made rigorous and precise using robust control theory (see below for details). Here, we will summarize the results and explain their significance for constraints that deconstrain. The idea in Fig. S5 is that there are hard constraints that depend on the structure of the interconnection (up vs. down and eyes open vs. closed). To make Fig. S5 quantitative, fragility is defined in terms of the net sensitivity of the closed-loop control system (see below). These constraints are in the form of tradeoffs between the fragility of the controlled system and the length  $l$ . The tradeoffs plotted in Fig. S5, however, apply only to the unavoidable part of the fragility that depends just on the dynamics of the pendulum and the measurement and actuation point. This portion of fragility is independent of any additional uncertainty (e.g., noise in nerves and muscles) that arises in the controller implementation. Robust control design must treat both sources of uncertainty, but this use of robust control provides additional insights that we expect will be crucial to understanding robustness in biology.

The theory then predicts that, with eyes closed, the up system is far too fragile to be stabilized by humans, whereas the down case is trivially stable without control, although there are still lower limits on achievable fragility in response to external disturbances. This theory also predicts that, for up/open and a sufficiently short length  $l$ , the system is also too fragile to be stabilized. Although exact prediction of that length depends on the noise and time delays within the human controller, the qualitative dependence summarized in Fig. S5 of the unavoidable fragility on up/down, open/closed, and length  $l$  allows for a variety of conclusions to be rigorously inferred. One conclusion is the obvious benefit to this control task of remotely sensing the tip using vision rather than relying on hand contact alone, independent of additional details. This perhaps intuitively obvious observation can, thus, be elegantly and rigorously formalized.

Another interesting prediction of the theory that can be tested experimentally is that the hand will tend to oscillate at lengths near the limit of control. There is no purpose to these oscillations per se, and they are simply side effects of the hard constraints on robust control (in contrast to myriad oscillatory phenomena with obvious benefits such as clocks, wheels, pistons, cell cycles, radio carriers, etc.) Furthermore, this increase in control effort observed as  $l$  is shortened is a universal telltale sign of system approaching breakdown, although in other systems, it may be manifested as loss of control variability because of actuator saturations. Similarly, the illusions described in the paper may similarly be side effects of the kind of robust control, revealing not simple flaws but the consequences of intrinsic tradeoffs when systems are pushed near limits.

Another immediate consequence of robust control theory is that the controller parameterization, no matter what the implementation substrate (human or robot), necessarily is large/thin as well as nonconvex. The technical details of what constitutes large/thin/nonconvex parameterizations are not simple, but the intuition is simple. The set of controllers that robustly stabilizes the up/open case is a vast and essentially infinite set of dynamical systems, but it is a vanishingly small subset of all controllers (just as with outfits vs. heaps). In other words, random controllers will almost surely not stabilize. Tuning parameters in a robotic controller without computer-aided design tools is essentially im-

possible. Furthermore, if two parameter sets each are robustly stabilizing, their mean may not be stabilizing, and this nonconvexity is another universal property of robust controllers. Indeed, this information is all very well-known, because a major element of robust control theory is constructing embeddings of controller parameters in higher-dimensional abstracted parameter spaces that are convex and thus, searchable. Thus, the (non) convexity of parameter spaces has been a subject of intense study and is entirely consistent with the other case studies in the paper.

## Pendulum Details

The pendulum example in Fig. S4 would correspond in Fig. S1 to  $r = 0$  and  $A$  modeling the combination of hand and pendulum, with  $C$  being the controller, including the eyes, brain, nervous system, and muscle actuator. The noise  $n$  is assumed to model all of the internal noises in the controller as well as any external disturbances (e.g., air currents) in the environment, but we will not aim for a detailed characterization. All of the signals are now functions of time, and  $A$  and  $C$  have dynamics that will be explored in more detail later. We will consider a simple model that is standard in undergraduate courses and laboratories, where the cart moves in a line and the pendulum moves in a plane. This example ignores all of the 3D motion that makes the problem only moderately more difficult to control but vastly more difficult to explain in any detail. This problem is seemingly unavoidable. Algorithms that underlie robust control scale well (polynomially) with problem size but are rarely solvable analytically. The kind of analytic theory that we will do here, helpful for gaining insight into the nature of feedback and dynamics, is only possible with extremely simplified examples.

The standard equations of motion for a cart and pendulum in a plane are (Eqs. S7–S9)

$$\begin{aligned}(M + m)\ddot{x} + ml(\ddot{\theta} \cos \theta - \dot{\theta}^2 \sin \theta) &= u, \\ \ddot{x} \cos \theta + l\ddot{\theta} \pm g \sin \theta &= 0, \text{ and} \\ y &= x + l \sin \theta.\end{aligned}\quad [\text{S7–S9}]$$

The force of gravity is  $g$ . The up position equations correspond to the  $-g$  in the  $\pm g$  term, and if the whole diagram is flipped upside down (and the pendulum is grasped lightly by the hand), then the down equations are for the  $+g$  case. It is not easy in practice to actually confine motion to a plane; however, this simplification will not hinder our use of this model, because we will be proving hard bounds on achievable robustness, and full 3D motion would simply make things more difficult. The pendulum stick is assumed to lack mass, with all of the mass concentrated at the tip.

The system in Eqs. S7–S9 can be linearized to get the transfer functions from the control input  $u(t)$  to both the output  $y(t)$  and the hand position  $x(t)$  [with transforms  $U(s)$ ,  $Y(s)$ , and  $X(s)$ ]. If we denote the transfer functions from  $U(s)$  to  $Y(s)$  and  $X(s)$  as  $A_Y(s)$  and  $A_X(s)$ , then we can solve analytically for (Eq. S10)

$$A_X(s) = \frac{ls^2 \pm g}{D(s)} \quad A_Y(s) = \frac{\pm g}{D(s)} \quad D(s) = s^2 [Ml s^2 \pm (M + m)g], \quad [\text{S10}]$$

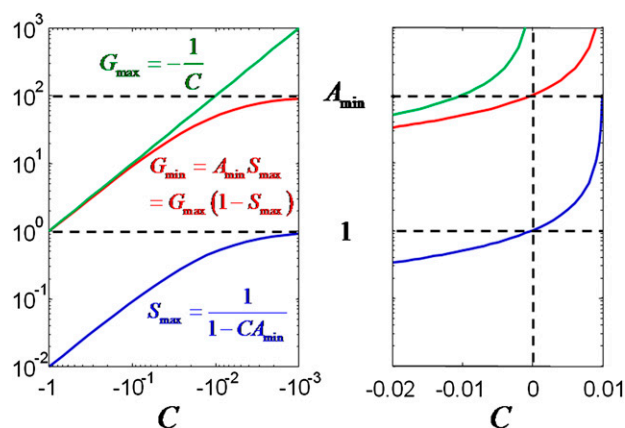
where  $-$  and  $+$  are for up and down, respectively. Note that, in the up case, both transfer functions are unstable, with a pole  $P > 0$  that solves  $D(p) = 0$ , and the  $X$  transfer function has a zero  $z > 0$  with (Eq. S11)

$$p = z\sqrt{1+r} = \sqrt{\frac{g}{l}(1+r)}, \quad z = \sqrt{\frac{g}{l}}, \quad r = \frac{m}{M}. \quad [\text{S11}]$$

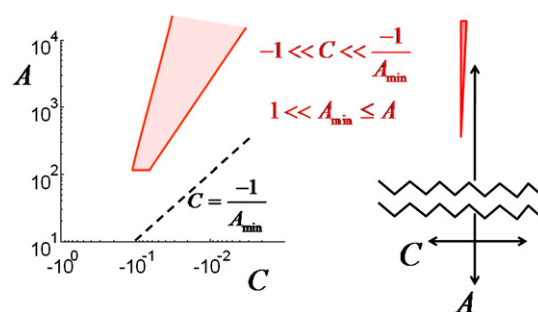
A standard control theory result is that the sensitivity function  $S(s)$  in Eq. S4 has the property that, if  $C$  is stabilizing, then



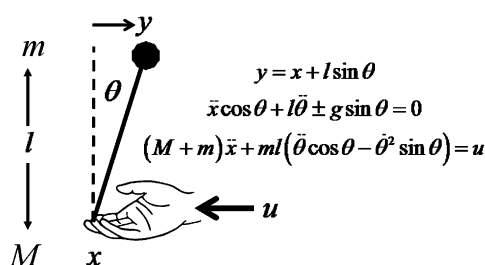




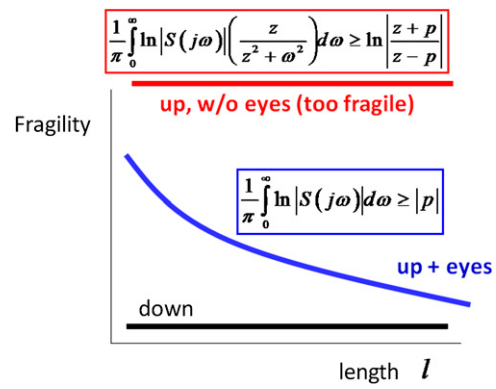
**Fig. S2.** System performance  $G_{\max}$  (green),  $G_{\min}$  (red), and  $S_{\max}$  (blue) as a function of component values  $C$  for  $A_{\min} = 100$ . *Left* has logarithmic units for  $C$ . In *Right*, note that, for positive feedback  $C > 0$ ,  $S_{\max} > 1$  and noise is amplified.



**Fig. S3.** Plot of the thin sliver  $1 \ll A_{\min} \leq A$  and  $-1 \ll C \ll \frac{-1}{A_{\min}}$  of functional parameter values for  $A$  and  $C$  in the space of all parameter values. *Left* is a log-log plot of a region blown up around the functional parameter values in red. *Right* is linear scaling, also focusing on the (red) functional region.



**Fig. S4.** Cartoon of the idealized stabilization problem with mass  $m$  at location  $y$  on top of an inverted (massless) pendulum of length  $l$  controlled by a control force  $u$  acting on the hand at position  $x$  and assumed to have effective mass  $M$ . The force of gravity is  $g$ . The up position equations correspond to the  $-g$  in the  $\pm g$  term, and if the whole diagram is flipped upside down (and the pendulum is grasped lightly by the hand), then the equations are for the  $+g$  case. The best experimental results are obtained with pendulums of widely varying lengths and masses, but they can also be performed with a standard mechanical extendable pointer (details in the text).



**Fig. S5.** Plots of hard constraint showing the net fragility as a function of the length of the pendulum and the structure of the controlled system. These cartoons can be made to be precise.